Wikipedia Administrative Pages Analytics

Marc Miquel i Ribé, PhD

User:marcmiquel / Catalan Wikipedia

marcmiquel@gmail.com
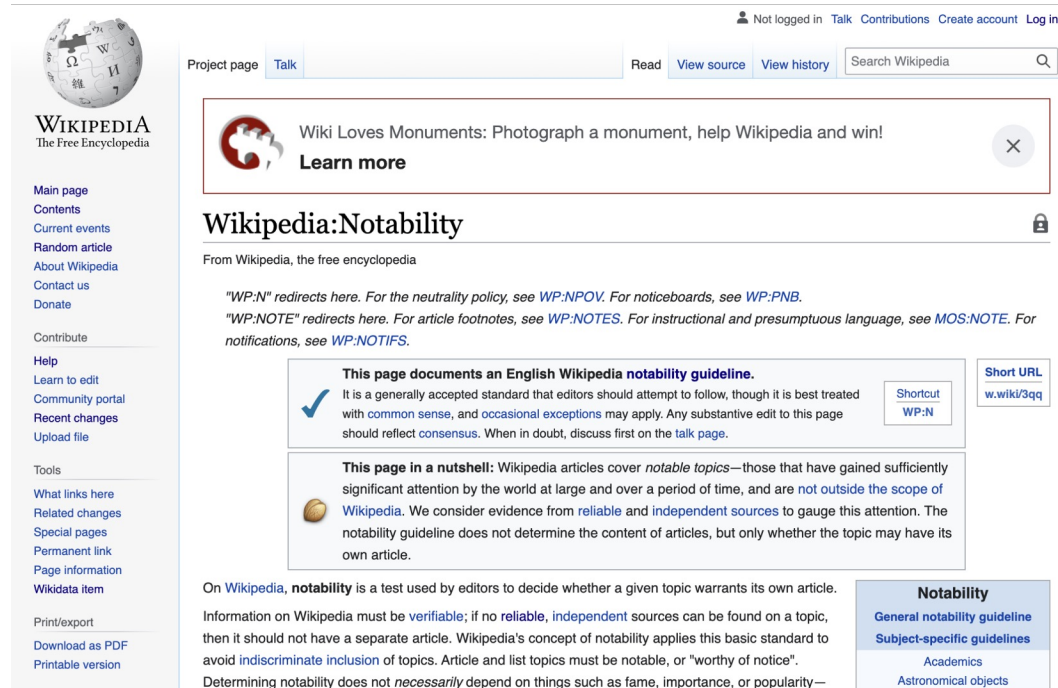
February 2023

Hello! My name is Marc Miquel,
I'm a Wikimedia researcher.

Wikipedia Administrative Pages.

1. Introduction

These are admin pages and they were created <u>twenty years ago</u>.

They are the backbone: what has enabled us to make Wikipedia successful.

2. Motivation

Welcome to the Jungle. **We need to research this.**

For Newcomers. Then, for all Wikipedians.

https://en.wikipedia.org/wiki/Wikipedia:Administration
https://en.wikipedia.org/wiki/Category:Wikipedia_administration

Image author: James Martins

Little Research on Admin Pages

1. Numerous and abundant rules and policies have been created that serve a wide variety of functions to maintain Wikipedia (Butler and Pike, 2008).

2. Newcomers find it difficult to have their contributions to policies accepted with an impact on their retention (Halfaker et al. 2013).

3. Participants (newcomers) to Wikipedia edit-a-thons experience moments of frustration when navigating a complex bureaucracy of policies and procedures (Gluza et al. 2021).

4. Similar rule-making activity across the five communities replicates and extends prior work on English language Wikipedia alone (Hwang and Shaw, 2022).

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/ Research_on_Wikipedia_Administrative_Pages

Little Research on Admin Pages

Wikipedia is one of the most researched objects. With studies about many aspects, which could also be explored for admin pages:

- **Topical analysis** (Halavais and Lackaff, 2008; Kittur et al. 2009).

Content gaps (Miquel-Ribé and Laniado, 2021; Redi et al., 2020).

- **Content biases and diversity** (Callahan and Herring, 2011; Beytía and Wagner, 2022)

- **Contextualization and cross-language analysis** (Hecht and Gergle, 2010; Hecht, 2013, Samoilenko et al. 2016).

- **Content quality models** (Lewoniewski et al. 2019; Halfaker and Geiger, 2020).

- **Content readability and style** (Setia et al. 2021, Liu et al. 2021).

- **Participation, and inequality** (Ortega and Robles, 2008; Shaw and Hargittai, 2018).

- **Cooperation and peer-to-peer communication** (Wilkinson and Huberman, 2007; Kittur and Krau, 2008).

- **Readership** (Warncke-Wang et al. 2015; Lehmann et al. 2014).

- **Vandalism** (Shachaf and Hara, 2010; Adler et al. 2011).

Admin pages are the "forgotten knowledge gap".

Hence, there is no framework or application to analyze the admin pages.
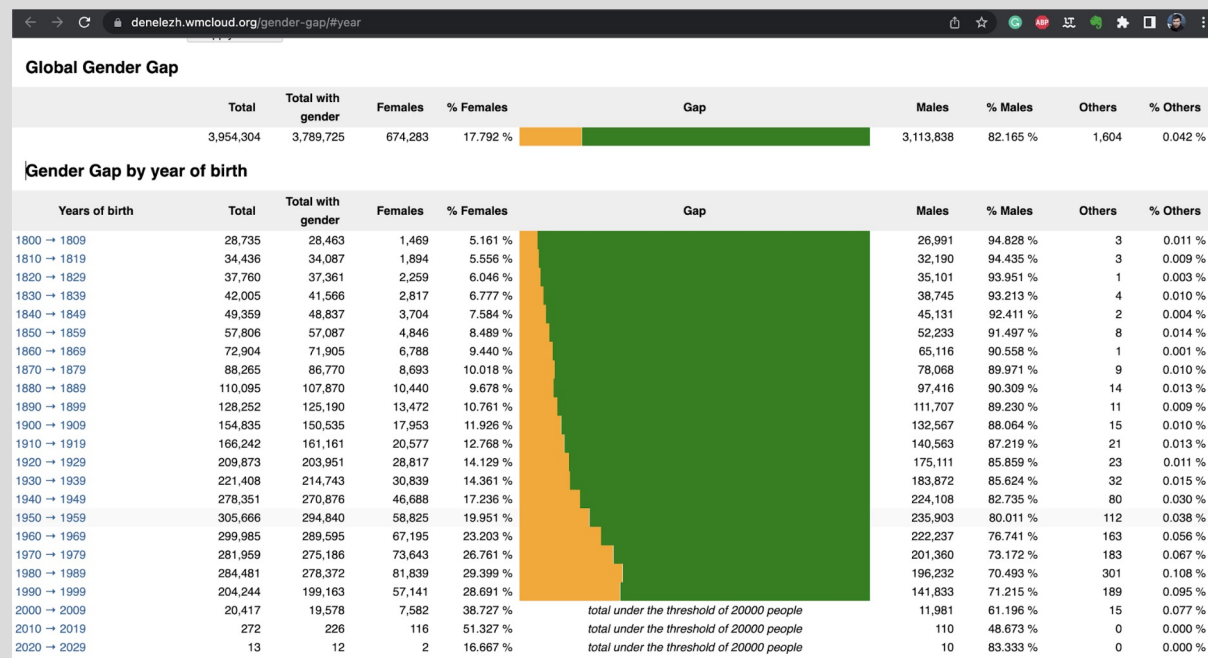
Gender Gap Dashboard (Denelezh, 2022)
https://denelezh.wmcloud.org/gender-gap/#year

Wikipedia Administrative Pages Analytics (WAPA) is a research project to measure and improve maintenance and inclusion in Wikipedia administrative pages across Wikipedia language editions.

"We envision a community that regularly consults analytical tools to make informed decisions on which pages to **create, improve or delete** *and, also, to modify the course of current processes in order to be* **inclusive** *to all Wikimedians."*
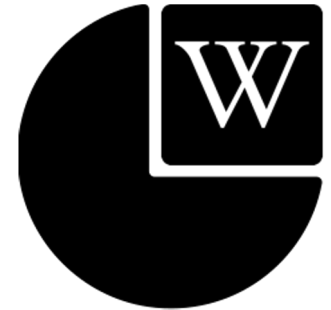
Project Principles: efficiency, maintenance and inclusion.

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics

2. Project Philosophy

"Let chaos (disorganized admin pages) reign,
and in chaos (with analytics) I will reign."*

Andrew Grove

* parentheses mine

3. Objectives

Objectives

1. Understand the state of development of the admin pages.

2. Classify the admin pages according to different types.

3. Create applications in order to support editing the admin pages.

4. Approach

In order to reach the three objectives, we need to:

1. First, collect the admin pages and analyze them (5.1).

2. Then, classify them into different types of pages (5.2).

3. Finally, create web applications with tools (5.3).

We are using the dumps provided at dumps.wikimedia.org.

The data processing and hosting of the applications happens at Toolforge.

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/Methodology
https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/Technical_Documentation
https://github.com/marcmiquel/WAPA

5. Results
5.1 Collecting the admin pages

What is an admin page?

Administrative Namespaces (No content or discussion oriented)

"Wikipedia requires a certain amount of administration and governance in order to further the project's goals.

To achieve Wikipedia's purpose, a wide range of **administrative pages** are made available in various namespaces which enumerate the various protocols and conventions created and implemented by community consensus for English Wikipedia."

https://en.wikipedia.org/wiki/Wikipedia:Administration

Wikipedia data structure			
Namespaces			
Subject namespaces		**Talk namespaces**	
0	(Main/Article)	Talk	1
2	User	User talk	3
4	Wikipedia	Wikipedia talk	5
6	File	File talk	7
8	MediaWiki	MediaWiki talk	9
10	Template	Template talk	11
12	Help	Help talk	13
14	Category	Category talk	15
100	Portal	Portal talk	101
118	Draft	Draft talk	119
710	TimedText	TimedText talk	711
828	Module	Module talk	829

Administrative Namespaces (No content or discussion oriented)
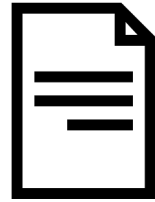
Template (ns10)
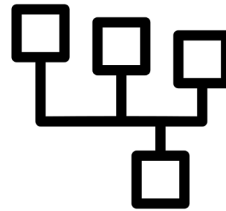
Categories (ns14)

Portals (ns100)

Wikipedia (ns4)

Help (ns12)

These two namespaces are the admin pages.

Admin Pages

| Language	Admin P. (%)	Content (%)	Admin Pages	Content	Sum
English	17,19	82,81	1,360,647	6,556,492	7,917,139
German	3,76	96,24	106,624	2,729,665	2,836,289
Swedish	1,09	98,91	28,091	2,553,869	2,581,960
French	2,34	97,66	58,892	2,459,001	2,517,893
Russian	3,91	96,09	75,661	1,857,186	1,932,847
Spanish	2,04	97,96	37,615	1,806,193	1,843,808
Italian	8,07	91,93	155,628	1,773,667	1,929,295
Polish	5,78	94,22	94,264	1,537,322	1,631,586
Japanese	9,88	90,12	147,390	1,345,075	1,492,465
Arabic	3,06	96,94	37,530	1,188,062	1,225,592
Portuguese	9,78	90,22	118,805	1,095,400	1,214,205
Catalan	1,18	98,82	8,499	711,831	720,330
Romanian	4,62	95,38	20,997	433,727	454,724
Armenian	2,51	97,49	7,563	293,417	300,980
Croatian	3,35	96,65	7,384	212,935	220,319
Yoruba	2,34	97,66	763	31,860	32,623
Hausa	0,94	99,06	179	18,815	18,994
Igbo	1,15	98,85	126	10,858	10,984
Twi	1,56	98,44	34	2,146	2,180

The proportion of admin pages in relation to all articles across languages ranges between 1 to 17%. Smaller Wikis have a smaller percentage.

* as of August 2022
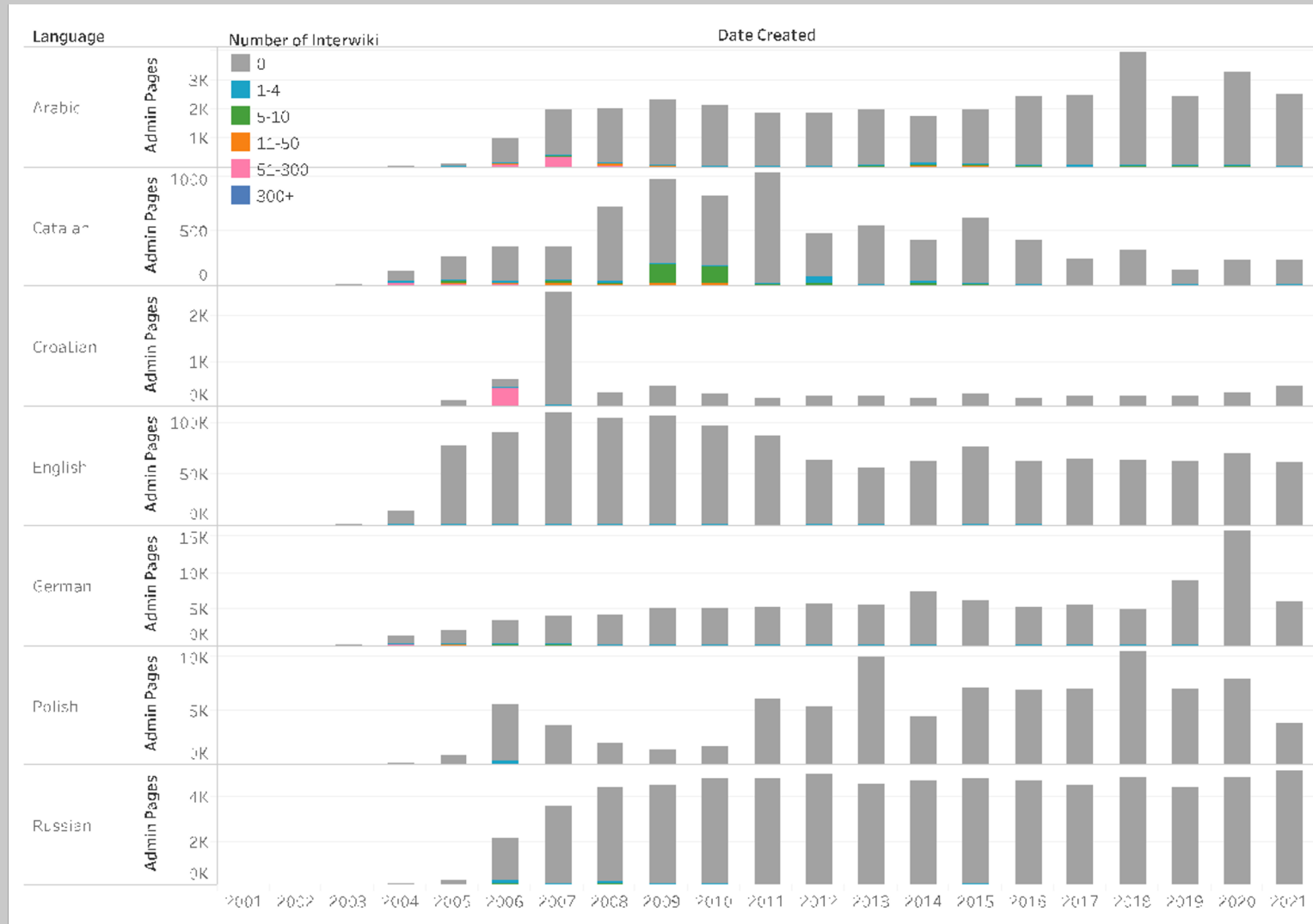
Creation of admin pages over time

There is a **constant creation** of admin pages over time. The majority are pages with no interwiki.

Creation of *valuable* admin pages over time

The creation of valuable admin pages (> 1 interwiki link) over time across languages. Different number of admin pages = Content Gaps. **We have an opportunity for import-export across langs.**
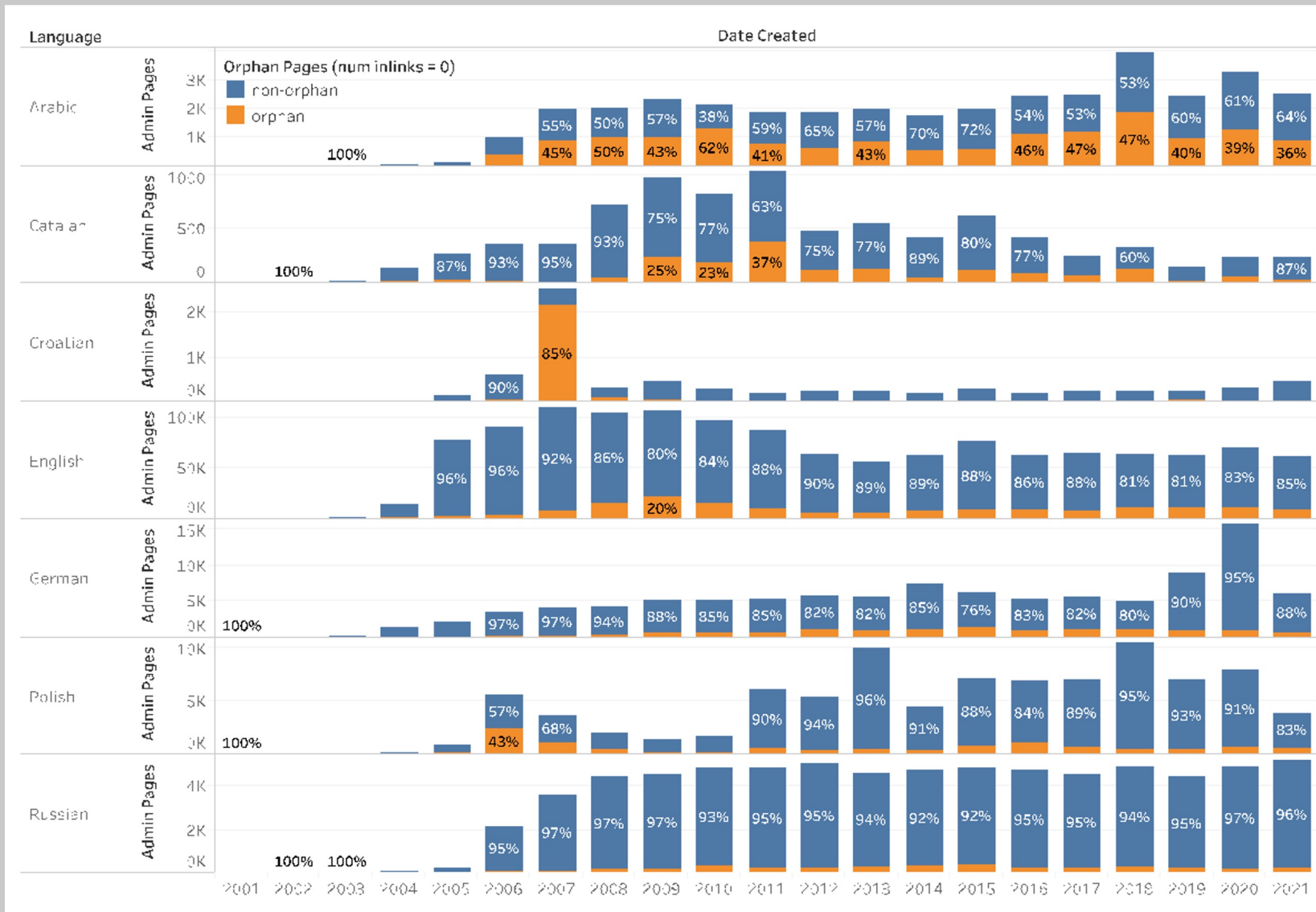
Creation of *orphan* admin pages over time

The creation of orphan (often non-valuable) admin pages over time across languages. Orphan = 0 inlinks.
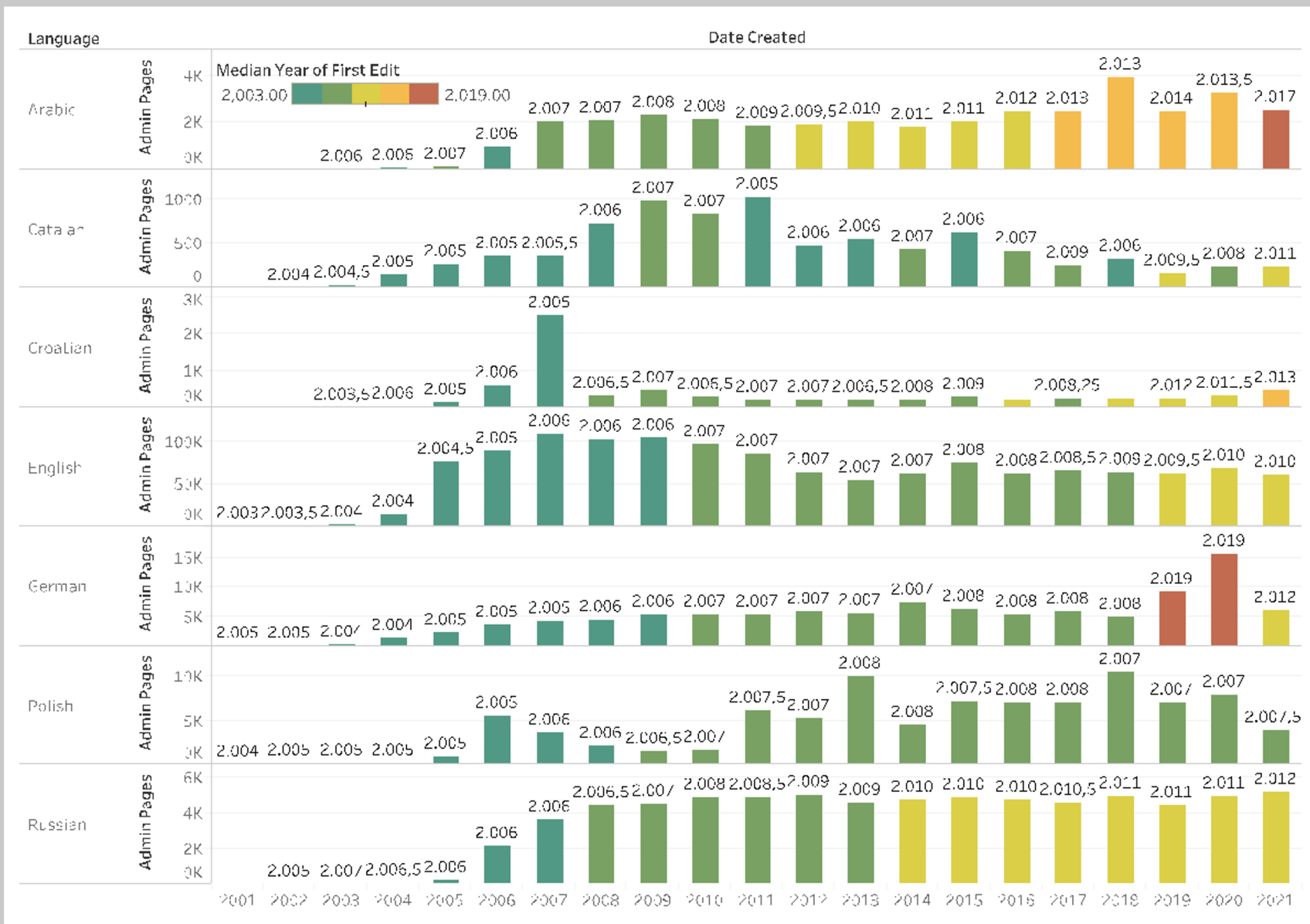
Creation of admin pages… by year of the first edit of their editors.
Created admin pages over time by the median year of the first edit of their editors (color).

Percentage of admin pages by days since the last edit.

In most languages, the majority of admin pages has not been edited once in 2 years.

Two conclusions stand out.

· **<u>Content Gaps</u>**
Very different number of valuable admin pages is available across languages, which means that there could be valuable content gaps, i.e., pages we could translate and help each other.

· **<u>Under-edited Pages</u>**
A small percentage of admin pages are edited in one year, which means that there are potential valuable non-edited pages that need more editing.

Editors of admin pages tend to be from previous generations. There is also an opportunity for inclusion.

5.2 Assigning admin pages to types

First, choosing the types:

We look for categories:
- More interwiki
- More subcategories/pages

Serving different purposes.

6 main types

Main categories within Wikipedia Administration

Main Types of Admin. Pages

The admin pages are in these two namespaces
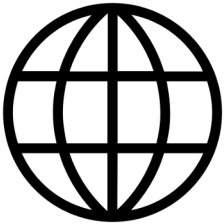
Wikipedia
(ns4)

Help
(ns12)

Policies & Guidelines

Help Pages

Essays

Village Pump

Tools

Wikiprojects

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/Methodology
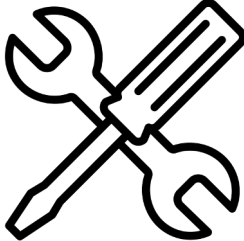
We assigned admin pages to types using the namespace, Wikidata properties and the category graph.

We selected the categories for the "category graph" that were the largest and with the most interwiki.

- Help Pages (ns12, wikidata property instance of: help page, category crawling)
- Policies & Guidelines (ns4, wikidata instance of, category crawling)
- Wikiprojects (ns4, wikidata instance of, category crawling)
- Essays (ns4, category crawling)
- Village Pump (ns4, category crawling)
- Tools (ns4, category crawling)
- Disclaimers (ns4, category crawling)
- Copyright (ns4, category crawling)

Often, one page belonged to multiple categories (e.g., essay and help page, policy and help page, etc.).
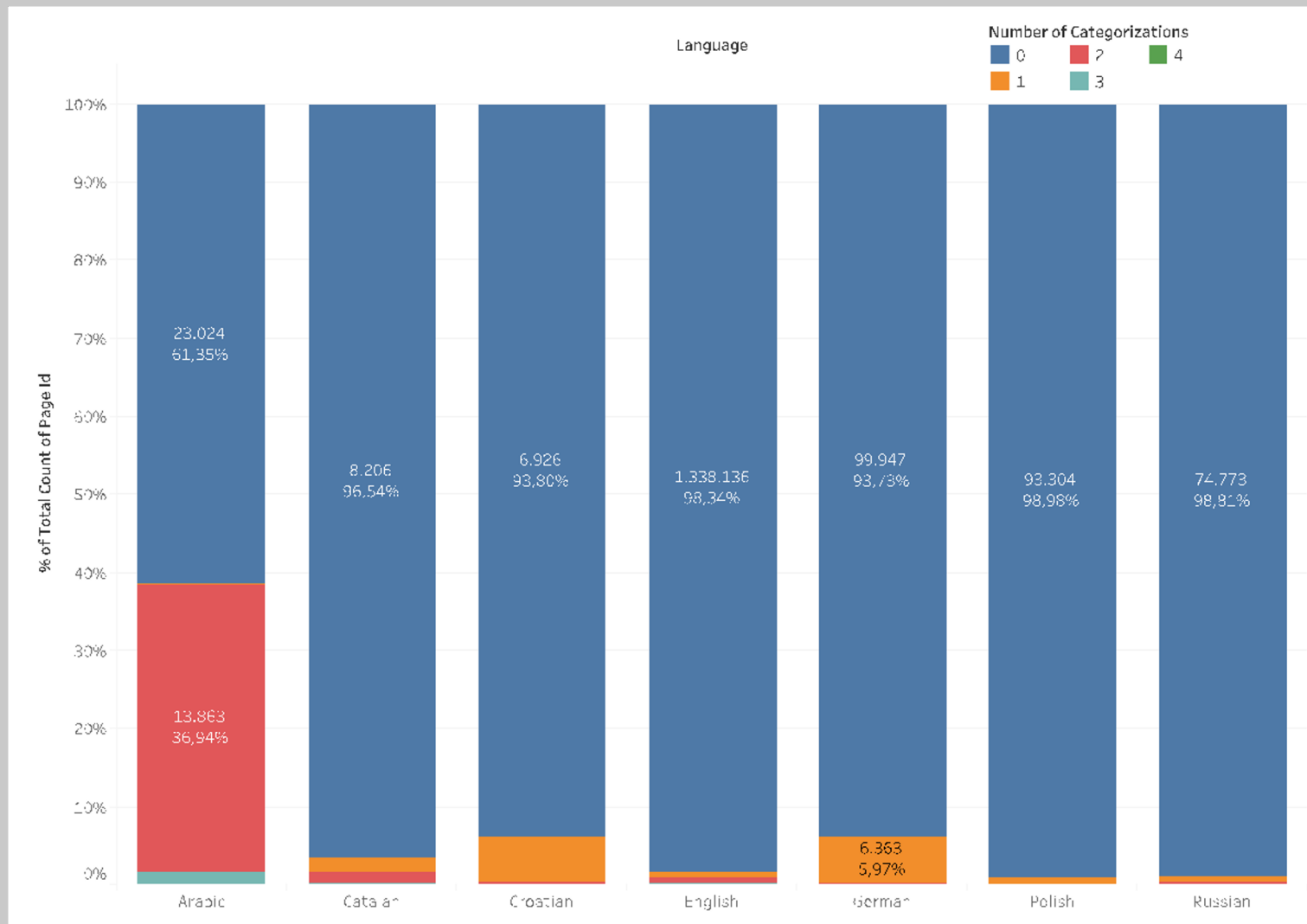
Most admin pages were not assigned.

Only a very small percentage of admin pages were identified through the methods exposed. We see that on average, it is below 5%.

5.3 Applications

How can we use the analysis of admin pages to stimulate editing?
Helping editors find *valuable* pages.

Screening tools

https://en.wikipedia.org/wiki/Screening_(medicine)

What is a valuable and a non valuable admin page?

- ***Completeness metrics:*** Number of Bytes, Number of References, Number of Images, Number of Outlinks.
- ***Relevance* metrics:*** Number of Inlinks, Number of Interwiki, Number of Multilingual Sister Projects.
- ***Popularity metrics:*** Number of pageviews.

- **Activity**. Number of Edits, Number of Edits last month, Number of Anonymous Edits, Number of Bot Edits, Active months, Editing days.
- **Disagreement**. Number of Discussions, Number of Reverts.
- **Engagement**. Number of Editors.
- **Inclusion**. Number of admin editors, Number of Edits last month by newcomer 60d, Number of Edits last month by admin, Number of Edits last month by anonymous, Median year first edit, Median editors Edits.
- **Recency**. Days last 50 edits, Days last 5 edits, Days last edit, Date of last discussion, Date of last edit. Number of edits last month.
- **Regularity**. Percent active months, Percent active days, Max. Active Months Row, Max Inactive Months Row.
- **Creation**. Date of creation, Language of the first timestamp, Total months.

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/Metrics

We created some **"screening tools"** 🔍
to address different purposes.

a) Page Across Languages

b) Page Gaps

c) Under-edited Pages
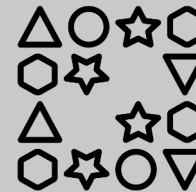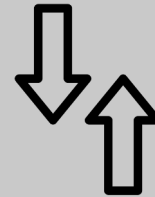
ALPHA VERSION

a) Page Across Languages

⇓⇑
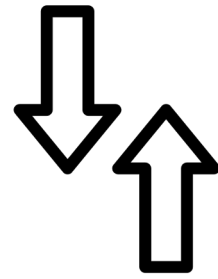
Introduce the page title and the metrics you want. Then you can compare the version of the page across languages.

Browse the results and sort the columns.

"Neutral Point of View" Page Characteristics Across Wikipedias

We choose:

- **Language + Article**
  or Qitem

- **Then the columns (e.g.,):**
    - **Date of creation**
    - **Days since last edit**
    - **Number of editors**
    - **Number of Bytes**
    - **Number of reverts**
    - **…**

https://wapa.wmcloud.org/page_across_languages

"Neutral Point of View" Page Characteristics Across Wikipedias

		languagecode	page_id	page_namespace	Title	Date of Creation	Number of Bytes	Number of Pageviews	Number of Interwiki	Number of Editors	Days Since The Last 5 Edits
		filter data								>1	
×	☐	jawiki	270	4	中立的な観点	20030202063539	50534	1556	122	81	38
×	☐	ruwiki	8379	4	Нейтральная_точка_зрения	20040719194740	35779	3563	122	114	108
×	☐	frwiki	3243	4	Neutralité_de_point_de_vue	20020913135834	21658	2168	122	190	160
×	☐	itwiki	2987	4	Punto_di_vista_neutrale	20021218085625	6355	1491	122	82	174
×	☐	ptwiki	1034	4	Princípio_da_imparcialidade	20030801145040	13683	1033	122	104	308
×	☐	plwiki	3620956	4	Neutralny_punkt_widzenia	20020919131421	32189	583	122	80	333
×	☐	hrwiki	690614	4	Nepristrano_gledište	20050826094735	28858	180	122	28	371
×	☐	svwiki	1744	4	Neutral_synvinkel	20021125173138	21675	177	122	88	604
×	☐	hywiki	19693	4	Չեզոք_տեսակետ	20091025153748	2200	67	122	7	726
×	☐	dewiki	3542	4	Neutraler_Standpunkt	20011124160207	17144	906	122	290	867
×	☐	eswiki	2411	4	Punto_de_vista_neutral	20020723224232	11290	4095	122	105	882
×	☐	arwiki	12934	4	وجهة_النظر_المحايدة	20050615084124	7955	870	122	41	88
×	☐	cawiki	7170	4	Punt_de_vista_neutral	20040113235956	6370	120	122	24	3105
×	☐	rowiki	90	4	Punct_de_vedere_neutru	20030722001845	3674	171	122	15	3362
×	☐	igwiki	1948	4	Neutral_point_of_view	20070321023249	742	108	122	4	3439

Sortable Table

https://wapa.wmcloud.org/page_across_languages

b) Page Gaps
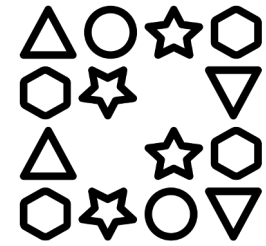
Search for admin pages in one language edition missing in another.
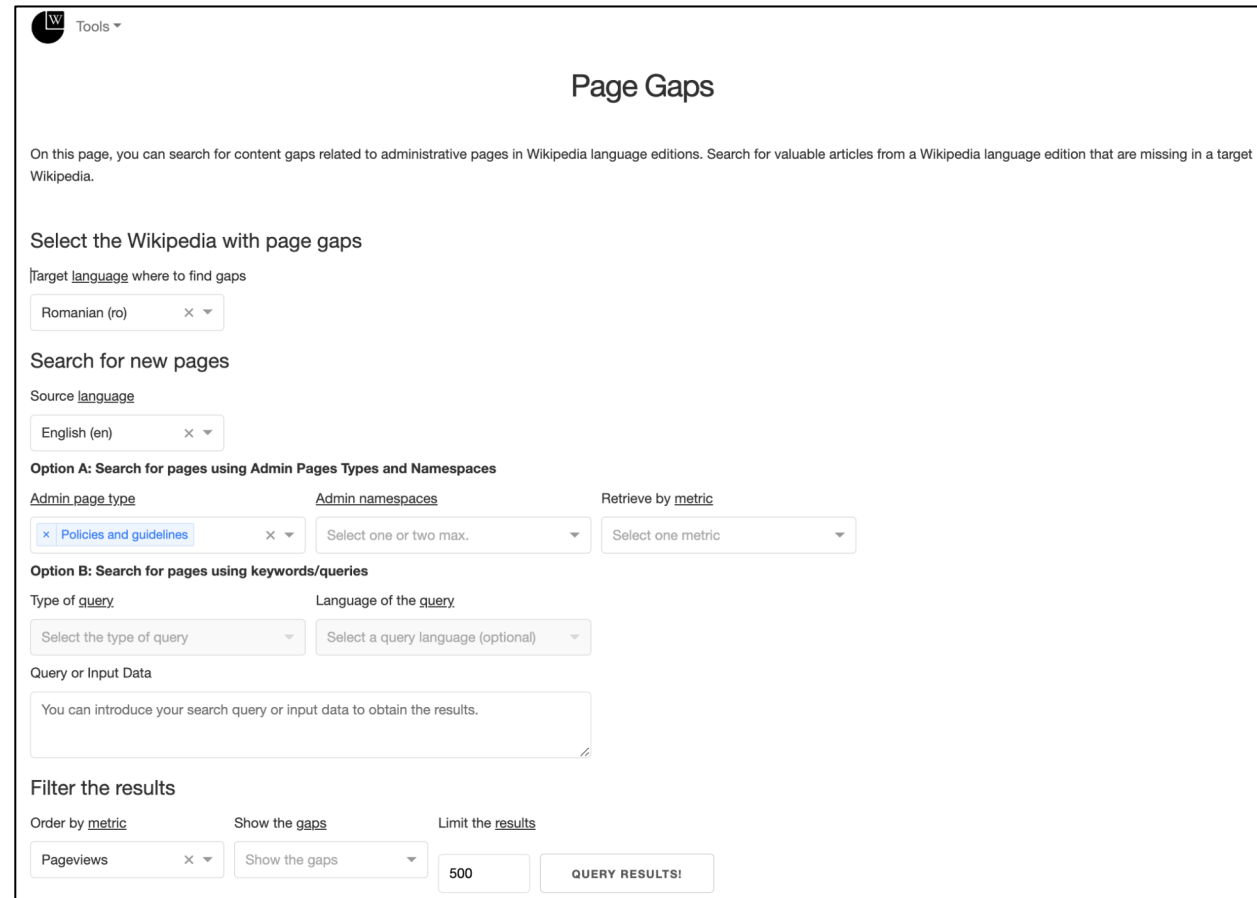
English Wikipedia Policies missing in Romanian Wikipedia

We choose:

- **Target Language (where gaps are)**

- **Source Language (where content is)**
- **Characteristics:**
  - **Admin page type**
  - **Admin namespaces**
  - **Retrieve by metric**

or search content (Wikidata, etc.).

- **Order the results by metric**

https://wapa.wmcloud.org/page_gaps/

English Wikipedia Policies missing in Romanian Wikipedia

N°	Qitem	English Title	Edits	Days Since The Last Edit	Interwiki	Inlinks	Editors	Pageviews	Romanian Title
1	Q10973440	Wikipedia:Project namespace	806	2	23	24619	267	41170	
2	Q5854231	Wikipedia:Village pump (proposals)	139440	1	47	44331	8954	16606	
3	Q10973854	Wikipedia:Writing better articles	1316	19	39	537891	546	11547	
4	Q10416758	Wikipedia:Naming conventions (technical restrictions)	1068	58	17	1964	369	10613	
5	Q22964187	Wikipedia:Wikipedia is not a reliable source	552	4	7	1887	147	10476	
6	Q10966628	Wikipedia:Manual of Style/Lead section	2402	14	32	111062	720	9879	
7	Q10818201	Wikipedia:Administrators' noticeboard/Edit warring	232746	1	9	17902	16105	9618	
8	Q5617729	Wikipedia:Accuracy dispute	844	61	14	12637	285	9080	
9	Q4657739	Wikipedia:Non-free content	2097	71	54	1362768	702	8373	
10	Q10990487	Wikipedia:Copyright violations	355	103	24	84759	168	8060	
11	Q13360396	Wikipedia:Appealing a block	346	6	11	739799	158	7515	
12	Q5641607	Wikipedia:Talk page guidelines	2222	2	36	985235	782	7436	
13	Q11051015	Wikipedia:Content disclaimer	139	169	46	1690	58	7306	
14	Q10823887	Wikipedia:Wikimedia sister projects	568	39	30	16104	201	6743	
15	Q7114571	Wikipedia:Village pump (idea lab)	27062	0	11	10757	2554	6691	
16	Q11107170	Wikipedia:Reviewing pending changes	383	19	25	4344	168	6344	
17	Q11015023	Wikipedia:Legal disclaimer	70	169	51	677	45	6340	
18	Q15100248	Wikipedia:Categorizing redirects	449	73	3	2141033	135	6128	
19	Q20978602	Wikipedia:List of controversial issues	2161	24	5	5060	636	5925	
20	Q13429672	Wikipedia:Requests for undeletion	111606	2	20	66992	20493	5754	

Valuable one (?)

Table with the gaps

https://wapa.wmcloud.org/page_gaps/

c) Underedited Pages

Find specific admin pages in Wikipedia language editions that are valuable, but might have not been edited enough, or for a long time.

Polish Wikipedia: Help Pages with more interwiki/pageviews and less edited.

We choose:

- **Language edition**

- **Characteristics:**
  - **Admin page type**
  - **Admin namespaces**

- **Retrieve by metric (2 = Pareto Front)**
  - **Interwiki**
  - **Pageviews**

- **Select other columns**
  - **Date of creation, Bytes, Inlinks…**

- **Order the results by metric**
  - **e.g., "Days Since the Last Edit"**

https://wapa.wmcloud.org/underedited_pages/

Polish Wikipedia: Help Pages with more interwiki/pageviews and less edited.

Pareto Front
- Yes
- No

Pageviews (y-axis)
Interwiki (x-axis)

Strona ujednoznaczniająca
Kawiarenka/Kwestie techniczne
Zadania dla botów
Strona użytkownika · VisualEditor
Nazwa użytkownika · Rurki
Zasady wstawiania Szablony · grafów zewnętrznych
WP
Śmiało modyfikuj strony
...enia grup użytkowników
Wikimedia Cloud Services
Komitet powitalny
Narzędzia
Nazewnictwo artykułów
Boty
Zalecenia dotyczące nazewnictwa biologicznego i systematyki
Dyskusja · Narzędzia/HotCat
Narzędzia cytowania
Standardy artykułów
Powiadomienia
Narzędzia/Navigation popups · grafiki
Linki
WikiURLskracacz
PetScan
Zgłoszenia · AutoWikiBrowser
Szablony/Dyskusja wikipedysty
Opisy licencji grafik
Narzędzia/refToolbar
VisualEditor/Dlaczego
Szablony/Ogólne · VisualEditor/Nawigacja
Powiadomienia/Podziękowania
Szablony/Źródła artykułów
VisualEditor/Opinie
VisualEditor/Strony pomocy · Narzędzia/LiveRC
Szablony/Wikipedysta
Narzędzia/wikEd
Huggle
Drzewo genealogiczne
Powiadomienia/Blogi · Szablony/Do usunięcia
Wikipedia i naukowcy
Kontrola autorytatywna
Narzędzia/EditHedaraphon
Szablony/Konserwacja
Narzędzia/wikEd/Pomoc
Szablony/Media
InputBox

Polish Wikipedia: Help Pages with more interwiki/pageviews and less edited.

Nº	Qitem	Page Title	Bytes	Days Since The Last Edit	Interwiki	Pageviews	Date of Creation	Edits	Editors	Inlinks	Pareto Front
1	Q10640405	WP:Opisy licencji grafiki	9.4k	861	52	28	2004-09-06	196	38	5780	Yes
2	Q13921062	WP:VisualEditor/Dlaczego	3.8k	633	8	15	2013-07-15	12	5	14	Yes
3	Q56723771	WP:PetScan	1.3k	626	14	25	2011-08-30	20	6	35	Yes
4	Q3396208	WP:Narzędzia/HotCat	1.9k	592	96	49	2008-04-15	40	11	58	Yes
5	Q3223292	WP:Szablony/Dyskusja wikipedysty	11.6k	455	36	22	2005-03-06	181	76	223	Yes
6	Q13666672	WP:VisualEditor/Nawigacja	1.2k	412	35	16	2013-07-10	25	7	29	Yes
7	Q10634368	WP:Najczęstsze nieporozumienia	8.7k	306	39	60	2002-11-25	93	36	25286	Yes
8	Q4048867	WP:Boty	16.8k	8	185	97	2003-03-04	468	152	649	Yes
9	Q6584911	WP:Narzędzia	22.2k	2	41	84	2006-08-01	158	54	1317	Yes
10	Q4167410	WP:Strona ujednoznaczniająca	12.6k	160	113	4925	2002-11-23	711	141	190878	No
11	Q4582194	WP:Kawiarenka/Kwestie techniczne	85.9k	1	76	3275	2006-04-14	41846	1101	3440	No
12	Q11168190	WP:VisualEditor	8.3k	87	145	597	2013-04-27	173	50	0	No
13	Q4592334	WP:Strona użytkownika	14.0k	265	97	571	2007-02-24	175	59	0	No
14	Q3186256	WP:Nazewnictwo artykułów	9.2k	433	101	82	2013-05-18	58	23	20201	No
15	Q4063270	WP:AutoWikiBrowser	2.6k	361	102	23	2006-09-15	58	20	63	No
16	Q4582561	WP:Zadania dla botów	21.1k	1	52	1549	2005-05-12	20134	596	649	No
17	Q4664077	WP:Nazwa użytkownika	4.8k	207	81	557	2005-11-27	196	79	0	No
18	Q4615128	WP:Boty/Zgłoszenia	2.9k	623	91	25	2006-09-30	418	88	322	No
19	Q3916099	WP:Śmiało modyfikuj strony	2.1k	2	84	205	2002-11-28	244	93	234302	No
20	Q28521	WP:Łasica	3.4k	450	80	24	2007-02-11	44	19	205	No

Valuable one (?)

Table sorted by days since the last edit

Conclusions

Take aways

- Support small wikis grow their *relevant* admin pages (Page Gaps)

- Keep updated/improve the popular admin pages and be inclusive inviting new editors (Underedited Pages).

- Admin pages are essential; they should be a bridge between Wikipedians, between generations.

- Visualizations can help in getting a glimpse of how well we are doing it.

- Analytical tools do not necessarily tell you to delete, but they can help you focus.

What can you do as **CEE communities**?

Editors, keep on editing. Affiliates: Please, create programs and get organized around admin pages.

We drew the ideas for the tools from conversations with fellow Wikimedians to identify "use cases" to generate tools.

Please, talk out your ideas: we need to identify and polish the use cases. Maybe there are better tools to make. All feedback is valuable!

Looking forward to Wikimania 2023.

What can we do as **researchers**?

- Improving the dashboards so that they encompass more use cases.

- Make available all the datasets and encourage research on admin pages.

- Bridges with product in order to help in the use of data.

- Other research lines: aspects related to readability.

If you have any **question**, you can e-mail me at
marcmiquel@gmail.com or Username:marcmiquel
or leave me a message at the project page.

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics

Project Mottos

Admin pages with great power, come with great maintenance.
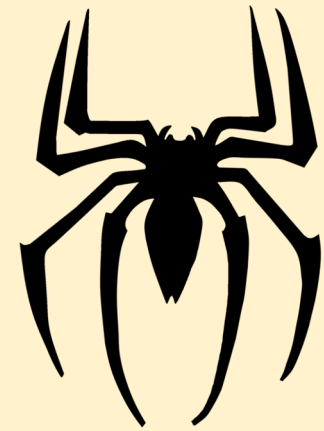
If it is everybody's Wikipedia, it must be everybody's admin pages.

Thank you very much!

Marc Miquel i Ribé, PhD

User:marcmiquel / Catalan Wikipedia

marcmiquel@gmail.com

WIKIMEDIA MEETING 2022